DOCUMENT RESUME

ED 046 979 TM 000 340

AUTHOR Busse, Thomas V.: And Others
TITLE Testing Conditions and the Measurement of Creative Abilities in Lower-Class Preschool Children.
PUB DATE Feb 71
NOTE 42p.: Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Creative Ability, Creativity Research, *Creativity Tests, *Lower Class, *Preschool Children, Preschool Tests, Racial Differences, Reinforcement, Sex Differences, *Test Reliability
IDENTIFIERS Construction Test, Pattern Meanings Test, Starkweather Test, Torrance Unusual Uses Test

ABSTRACT

The effects of play-like, verbal-feedback, and nonverbal-feedback testing conditions on three creative ability measures and the reliabilities of those measures were studied using 175 lower-class preschool children. The creativity measures were found to be largely unaffected by variations in testing conditions; but at the same time, different forms of the measures hypothesized to be parallel proved to have only small relationships with each other. However, split-half reliabilities of the individual measures were fairly high. (Author)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ED046979

Testing Conditions and the Measurement of Creative Abilities in Lower-Class Preschool Children[1]

Thomas V. Busse, Patricia Blum, and Marilyn Gutride[2]

Temple University

Researchers seeking environmental determinants of creative behaviors have been restricted by the lack of research documenting the reliability and validity of measuring instruments available for use with preschool children (Busse, 1969; Starkweather, 1964). Moreover, there is some evidence to show that measures currently used to assess creativity are not at all relevant for preschool children (Foster, 1967; Ward, 1968, 1969; Wodtke, 1964).

There is also conflicting evidence concerning the effect of various testing conditions on the measurement of creativity. A study by Savoca (1965) suggests that some measures available for assessing creative abilities of preschool children are affected by testing conditions. He found that giving trinkets as a reward to four-year-old children produced significantly higher scores on three of Guilford's (1967) divergent abilities than did nonevaluative verbal encouragement. Savoca's results held across both race and social class; however, his

TM 000 340

results show no breakdown by sex. Ward, Kogan, and Pankove (1970) using lower-class Negro fifth-graders likewise found that the giving of pennies resulted in higher fluency scores. Differences in response to the reward were not reported by sex.

Wallach and Kogan (1965) implied that children achieve higher scores on creativity measures in situations that minimize testing and evaluative aspects and maximize a game- or play-like context. Rogers (1959) took a similar position.

Dentler and Mackler's (1964) research, showing that children scored higher on an unusual uses test under relaxed conditions than under evaluative conditions, supports Wallach and Kogan's theorizing. However, since the statistical analysis reported by them is unclear, the meaning of their findings is uncertain.

Wallach and Kogan's theorizing was also supported by the findings of Boersma and O'Bryan (1968). Using Torrance's creativity tests with fourth-grade Canadian boys, they found that a nonevaluational context produced higher scores than did a conventional school setting.

In contrast, Wallach and Kogan's theorizing was not supported by the studies of Kogan and Morgan (1969) and Williams and Fleming (1969). Kogan and Morgan studied the effects on creativity scores of test- and game-like contexts. No clear-cut differences were apparent between the two contexts. The effects that Kogan and Morgan found varied depending on the task and the sex of the subjects as well as several other factors not directly relevant to the study reported here. Williams and Fleming tested three and four-year-old white, middle-class children under evaluative and play conditions. Their results, showing that both visual and verbal associative fluency scores were higher in the evaluative condition, also failed to support Wallach and Kogan's position. Sex differences were not studied by Williams and Fleming.

The present research investigated both the effects of three test conditions (play-like, verbal feedback, nonverbal feedback) on selected measures of preschool children's creative abilities and the reliabilities of those measures.

Method

Subjects

Ninety four-year-old Negro and 90 four-year-old white children from disadvantaged urban areas were tested with Form A of several creativity tests. Each group was evenly divided by sex. All of the children were attending one of eight centers of the local Head Start program.[3] Five children could not be administered form B of the creativity tests. Four of these children could not be located, and one child refused to be tested. This left 175 children who were tested on both forms A and B of the creativity tests.

The median age of the 44 white boys was 4-8 with a range of 3-11 to 5-0 at the time of form A testing. The median age of the 44 white girls was 4-5 with a range of 3-10 to 5-1. Forty-three Negro boys had a median age of 4-9 with a range of 3-11 to 5-2. Forty-four Negro girls had a median age of 4-8 ranging from 4-1 to 5-3.

Creativity measures

Four creativity measures were pretested on forty children both to evaluate their suitability for use with lower-class preschool children and to obtain a sample listing of responses. The measures listed below were

chosen because of their prominence in past research with young children:

1. Construction Test (Savoca, 1965). In form A of this test the child was asked to build four structures with a set of blocks, one structure at a time. The child was then asked to name the block structure. This form is taken directly from Savoca's work; but the use of time limits and the photographing of completed structures were eliminated. The procedures were similar for form B except that a set of Tinker Toys was substituted for the blocks.

2. Starkweather Test (Starkweather, 1964). Each form consists of a set of forty styrofoam forms. The child was asked what each of the forms could be. Forms A and B in this research are the same as those used by Starkweather. The forms consist of ten shapes each of which is included four times in four different colors (See Figure 1).

- - - - - - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - - - - - -

3. Pattern Meanings Test (Wallach and Kogan, 1965; Ward, 1968). The child was asked to suggest different

meanings for a series of line drawings. Form A in the present research used the same eight figures as did Wallach and Kogan. Eight additional form B figures were created for this research (See Figure 2). The figures were drawn in a variety of colors.

- - - - - - - - - - - - - - - - -

Insert Figure 2 about here

- - - - - - - - - - - - - - - - -

4. Unusual Uses Test (Torrance, 1962; Ward, 1968). In the version of this test used, the child was asked to think of unusual and different uses for a set of objects individually presented to him. The objects included: a brick, a string, a tin can, a pencil, a toy dog, a toy monkey, a book, and a piece of paper.

The Unusual Uses Test was dropped after pretesting because it appeared that the children were not relating to it; few of the forty pretest children gave meaningful responses to the items.

The three remaining tests were always given in the following order: Construction Test, Starkweather Test,

Pattern Meanings Test. The two forms of the tests were administered about two months apart. Each test was scored for fluency, flexibility, and originality.

It was anticipated that the three scores from a single test might be highly intercorrelated. However, since many previous investigators (e.g., Torrance, 1962; Ward, 1968; Yamamoto and Chimbidis, 1966) used these or similar multiple scores, it seemed valuable to look at the effects of the testing conditions on each score separately while recognizing that scores from the same test were not independent.

The fluency score for a test was the number of appropriate responses (non-bizarre responses) the subject gave. All responses of the pretest subjects to a particular test were independently scored as appropriate or inappropriate by two judges. Differences in scoring were resolved by discussion between the judges. This master list of appropriate and inappropriate responses was memorized by the two judges who served as Es for the subsequent testing. In addition to the responses that

occurred in the pretest sample, the Es used their judgment on other responses given by the children in the final sample. It was thought that the pretest administration and scoring experience had given the Es a sufficient framework to judge most new responses. The Es could not wait until later to judge the appropriateness of the responses because two of the testing conditions (to be explained later) required immediate feedback to the subject. The above procedures were followed in scoring the appropriateness of responses to the Pattern Meanings Test and the Starkweather Test. However, for the Construction Test, E had to judge at the time of the testing whether a subject's name for a structure was appropriate. Duplicate responses to an item and "I don't know" responses were considered inappropriate.

The reliability of the appropriateness scoring was checked on a sample of thirty children for two representative tests. The Spearman rank-order inter-rater reliability for appropriateness scoring was .95 for the Starkweather Test (form A) and .99 for the Pattern Meanings Test (form A).

The flexibility score for a particular test was the number of different categories of response into which a subject's appropriate answers were scored. The categories were constructed using the pretest data. For example, the flexibility categories for the Construction Test, form A, included: transportation, toys, furniture, animals, structures, appliances, clothing, topographical features, persons, letters or numbers, plants,

Two raters scored each of the protocols of the 175 subjects for flexibility. Inter-rater Spearman rank-order correlations for the Construction Test, form A, was .89; for the Construction Test, form B, .93. For the Starkweather Test, form A, the inter-rater correlation was .99; and for the Starkweather Test, form B, .99. The Pattern Meanings Test showed inter-rater correlations for flexibility of .96 for form A and .94 for form B.

A subject's originality score for a particular test was the number of appropriate responses to a particular item that occurred no more than twice in the responses of the 175 children to that item. The criterion of two occurrences rather than one was used because of the relatively large number of subjects in this study. To obtain the originality score, all responses of the subjects were catalogued.

Research design

The experimental design was a 2 (races) x 2 (sexes) x 3 (testing conditions; play, verbal feedback, nonverbal feedback) multivariate analysis of variance. Fifteen subjects were assigned to each cell.

All of the subjects were separated by race and sex and then randomized into the testing conditions. Each subject was individually administered forms A and B of the three creativity measures under a single testing condition. In light of Rosenthal's (1966) findings of various experimenter effects, approximately one half of the subjects (eight and seven respectively) in each cell of the factorial design were randomly assigned to each of the female experimenters. Each subject received forms A and B from the same experimenter.

Testing conditions

Play. Evaluational aspects were minimized in this situation. The subjects were invited to play some "children's games." The testing procedures employed were similar to those of Wallach and Kogan (1965). The E consistently encouraged the child and responded favorably to all of the child's attempts at the tasks. No feedback as to the correctness of the children's responses was given.

For example, in the Construction Test, form A, E said: "Do you ever play with blocks? We are going to play a game with them now. You may use as many blocks as you wish to build something. When you have finished, tell me what you have built and then you may build something else. O.K.? Go ahead." In general, E endeavored to make the situation as play-like and non-evaluative as possible. There were no time limits stated or implied in this condition.

Verbal feedback. In this test condition the E responded with "Right" for each appropriate answer and "Wrong" for each inappropriate answer. As described previously, a list of appropriate and inappropriate responses was compiled from the pretests of 40 subjects. The E's familiarity with the tests allowed them to judge additional responses. However, if the appropriateness or inappropriateness of a subject's answer was uncertain to the E, she responded with a "What else?" statement rather than with positive or negative feedback.

The directions for the Construction Test, form A, are typical of those for this test condition: E said, "Here are some blocks with which you can build things. Make

something with them using as many blocks as you want. When you have finished tell me what you have built and I will tell you if what you have made is right or wrong. O.K.? Go ahead." E kept the situation as test-like as possible, but no time limits were stated or implied.

Nonverbal feedback. In this testing condition E gave the subject a token for each appropriate answer, and sounded a loud clicker after each inappropriate answer. The tokens were exchangeable for one of five prizes. In this test condition there was no verbal feedback to the subject's answers. For the Construction Test, form A, E said: "Here are some blocks with which you can build things. Make something with them using as many blocks as you want. When you have finished tell me what you have built. If you are right, I will give you one of these tokens. If you win enough of these tokens you can use them to buy one of these prizes (E shows the five different prizes to S). If your answer is wrong, I will click this clicker (E clicks clicker for S). O.K.? Make something with the blocks using as many as you want. Go ahead." As in the verbal feedback condition, E established a test-like situation, but no time limits were stated or implied.

Following testing, all children in the nonverbal condition were told that they had done very well and were given their choice of the prizes.

Results

Transformations

Three tests, the Construction Test, the Starkweather Test, and the Pattern Meanings Test were given to each subject. Each test yielded three scores: fluency, flexibility, and originality. Each test had two forms. Thus eighteen scores were obtained for each subject.

However, four scores could not be transformed to meet the requirements of a normal distribution. These four, Construction Test forms A and B -- fluency, and Construction Test forms A and B -- originality, were dropped from further analyses. The remaining fourteen variables, the transformations required to normalize them, and their raw score means and standard deviations are shown in Table 1.

- - - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - - -

Intercorrelations

The intercorrelations of the fourteen variables for 175 subjects are shown in Table 2. Ignoring for a moment the various factorial divisions within the data, it can be seen that the flexibility score of the Construction Test has an inter-form reliability (with approximately a two-

month interval between forms) of .17. The reliability coefficients reported here are thus composed of both inter-form and stability components. Similar reliabilities for the other six scores are .55 for Starkweather Test fluency, .55 for Starkweather Test flexibility, .45 for Starkweather Test originality, .25 for Pattern Meanings Test fluency, .32 for Pattern Meanings Test flexibility, and .23 for Pattern Meanings Test originality. Correlations of .15 and .20 are significant respectively at the five and one per cent levels.

- - - - - - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - - - - - -

Table 2 shows that the scores taken from the same measure are highly correlated. The three scores computed from the Starkweather Test have a minimum intercorrelation of .79 for form A and .80 for form B. The fluency and flexibility scores from the Pattern Meanings Test are likewise highly related (.93 for form A and .90 for form B). However, for the Pattern Meanings Test, the originality score is only moderately related to the fluency and flexibility scores (the correlations are .45 and .53 for form A, and .21 and .29 for form B).

The odd-even reliability coefficients corrected by the Spearman-Brown formula for a test of doubled length are shown in Table 3 for the Starkweather and Pattern Meanings tests (split-half reliabilities were not calculated for the Construction Test because of its very low scores). The highest reliabilities were obtained for the fluency scores and the lowest for the originality scores. The coefficients for forms A and B are of the same magnitude. All reliabilities except for the originality scores from the Pattern Meanings Test are adequate, if not exceptional.

- - - - - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - - - - -

Analyses of variance

The seven form A variables and the seven form B variables were then separately run in 2 (races) by 2 (sexes) by 3 (testing conditions) multivariate analyses of variance (Jones, 1966). Forms A and B were analyzed separately because of the possibility of form A sensitization effects on form B. In addition, because the B forms of the Construction and Pattern Meanings tests were specially constructed for this research, it was thought that the testing conditions might produce effects on them

different from those produced on the previously used form A versions. All factors were considered to be fixed. The calculations were done by Finn's (1968) computer program. The likelihood ratio criterion (lambda) was used to test the multivariate hypotheses.

Each of the cells originally contained fifteen subjects. However, five subjects could not be given form B of the tests. In order to allow for a simplified analysis, seven additional subjects were randomly dropped from appropriate cells in order to leave 14 subjects in each cell. Fortunately each of the five subjects who could not be retested came from different cells and had been tested by the E with eight subjects assigned to her in the particular cell in question. Thus, by randomly dropping additional subjects from the E assigned eight Ss in a particular cell, each cell was left with seven subjects tested by E1 and seven subjects tested by E2.

As can be seen in Table 4, none of the multivariate Fs for the seven form A scores were significant. The few univariate Fs that reach significance are apparently chance findings.

- - - - - - - - - - - - - - - -

Insert Table 4 about here

- - - - - - - - - - - - - - - -

As shown in Table 5, none of the multivariate Fs except for the sex by testing condition interaction were significant for the seven form B variables. The sex by testing condition interaction yielded a multivariate F of 2.47 ($p < .003$).

- - - - - - - - - - - - - - - -

Insert Table 5 about here

- - - - - - - - - - - - - - - -

The univariate Fs for the seven form B variables listed in Table 5 indicate that the three scores (fluency, flexibility, and originality) from the Pattern Meanings Test were the primary source of the significant multivariate F for sex by testing condition. Although only the univariate F for fluency reached the .05 level of significance, the univariate Fs for flexibility and originality both approached significance.

The relevant cell means for the Pattern Meanings Test, form B, fluency score are shown in Table 6. The cell means for the flexibility and originality scores are similar.

- - - - - - - - - - - - - - - - - -

Insert Table 6 about here

- - - - - - - - - - - - - - - - - -

Tukey's multiple comparison test was used to test the differences between means for the Pattern Meanings Test, form B, fluency score. It was found that none of of the boys' means for the three testing conditions were significantly different from one another. For girls however, the mean for the verbal feedback testing condition was significantly higher than that for the play condition ($p < .05$). In addition, the mean for boys in the play condition was significantly higher than for girls in the play condition ($p < .05$). There were no sex differences in either the verbal feedback or the nonverbal feedback testing conditions.

Discussion

Wallach and Kogan (1965) implied that when creativity measures are administered in a situation that maximizes the subject's freedom to respond and minimizes evaluational aspects of the procedure children achieve higher scores than when the measures are given in a testing or evaluational context. Research findings on this point, as detailed in the introduction, are mixed.

The present research indicated that the form A versions of the Construction Test, the Starkweather Test, and the Pattern Meanings Test and the form B versions of the Construction and Starkweather tests were not affected by the testing conditions.

However, the form B of the Pattern Meanings Test constructed specially for this research showed a sex by testing condition interaction. The results for boys for this test support Wallach and Kogan's assumption that the play condition is most beneficial. But contrary to Wallach and Kogan's ideas, girls in the play condition achieved lower scores than did girls in other conditions.

This interaction finding for form B of the Pattern Meanings Test supports the results of Kogan and Morgan (1969), who found that the sex of the children interacted with the testing conditions on various creativity measures. However, this significant interaction might also be related to the sole use of female experimenters.

It should be emphasized that most of the creativity measures showed no significant effects due to the test conditions. It seems reasonable to conclude that most measures of this kind when given to preschool children are not easily altered by test condition variations of the type used in this study.

The inter-form reliabilities for the scores from the Pattern Meanings Test and the Construction Test are well below acceptable standards. The reliabilities for the Starkweather Test were only moderate at best and certainly are not comparable to those considered acceptable for published tests. These low creativity test reliabilities for very young children are consistent with those found by Wodtke (1964). Somewhat in contrast, split-half reliabilities reached more acceptable levels except for the originality scores from the Pattern Meanings Test. Moreover, the fact that scores from the Starkweather Test showed different univariate score distributions for forms A and B suggests that they are not true parallel forms. Thus, it is probably inadvisable to use the Pattern Meanings Test or the Construction Test as a creativity measure for four-year-old children. Furthermore, the Starkweather Test should be used cautiously with this age group.

The intercorrelations between creativity scores from different tests were generally significant but very low. Thus it is doubtful that any unitary "creative ability" exists at this age. Low intercorrelations from other studies (Ward, 1968, 1969) support this view.

Most of the distributions of the creativity scores

used in this research were skewed. A part of this skewness can be attributed to a few subjects who obtained scores many standard deviations above the mean. For the most part different subjects were extremely high for different measures. The fact that these extreme scores occurred for seven out of fourteen variables suggests that some creative abilities may not be normally distributed in the population sampled here. Therefore, it is suggested that all data from measures of the type used in this study be checked for normality.

The fact that a few subjects scored many standard deviations above the mean also has larger implications. The generally accepted mental abilities model assumes an underlying normal distribution of abilities in the general population. But if a few children possess the abilities measured here to an extremely high degree, a "genius" model for these creative abilities might be more appropriate.

For some children in the nonverbal reinforcement condition, the clicker seemed to function as a positive rather than a negative reinforcer. Thus these children would have received, in effect, no negative reinforcement. Caution should be exercised in future studies in the use of such reinforcers.

In conclusion, the creativity measures used here were found to be largely unaffected by variations in testing conditions; but at the same time, different forms of the creativity measures hypothesized to be parallel proved to have only small relationships with each other. These findings indicate a need for the development and evaluation of reliable and valid creativity measures for young children. Only after such measures are developed can researchers proceed with longitudinal and cross-sectional studies of the development of creativity.

References

Boersma, F. J. and O'Bryan, K. An investigation of the relationship between creativity and intelligence under two conditions of testing. Journal of Personality, 1968, 36, 341-348.

Busse, T. V. Child-rearing antecedents of flexible thinking. Developmental Psychology, 1969, 1, 585-591.

Dentler, R. A. and Mackler, B. Originality: Some social and personal determinants. Behavioral Science, 1964, 9, 1-7.

Finn, J. D. Multivariance--Univariate and multivariate analysis of variance, covariance, and regression: A fortran IV program. (Version 4) Buffalo, New York: State University of New York at Buffalo, 1968.

Foster, J. T. Some effects of preschool experience on the general intelligence and creativity of culturally disadvantaged children. (Doctoral dissertation, University of Southern California) Ann Arbor, Michigan: University Microfilms, 1967. No. 67-403.

Guilford, J. P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.

Jones, L. V. Analysis of variance in its multivariate developments. In R. B. Cattell (Ed.), Handbook of multivariate experimental psychology. Chicago: Rand McNally, 1966. Pp. 244-266.

Kogan, N. and Morgan, F. T. Task and motivational influences on the assessment of creative and intellective ability in children. Genetic Psychology Monographs, 1969, 80, 91-127.

Rogers, C. R. Toward a theory of creativity. In H. H. Anderson (Ed.), Creativity and its cultivation. New York: Harper and Row, 1959. Pp. 69-82.

Rosenthal, R. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966.

Savoca, A. F. The effects of reward, race, IQ, and socio-economic status on creative production of preschool children. (Doctoral dissertation, Louisiana State University) Ann Arbor, Michigan: University Microfilms, 1965. No. 65-11,405.

Starkweather, E. K. Problems in the measurement of creativity in preschool children. Journal of Educational Measurement, 1964, 1, 109-113.

Torrance, E. P. Guiding creative talent. Englewood Cliffs, New Jersey: Prentice-Hall, 1962.

Wallach, M. A. and Kogan, N. Modes of thinking in young children. New York: Holt, Rinehart and Winston, 1965.

Ward, W. C. Creativity in young children. Child Development, 1968, 39, 737-754.

Ward, W. C. Creativity and environmental cues in nursery school children. Developmental Psychology, 1969, 1, 543-547.

Ward, W. C., Kogan, N. and Pankove, E. Motivation and ability in children's creativity. Proceedings, 78th Annual Convention, American Psychological Association, 1970, 285-286.

Williams, T. M. and Fleming, J. W. Methodological study of the relationship between associative fluency and intelligence. Developmental Psychology, 1969, 1, 155-162.

Wodtke, K. H. Some data on the reliability and validity of creativity tests at the elementary school level. Educational and Psychological Measurement, 1964, 24, 399-408.

Yamamoto, K. and Chimbidis, M. E. Achievement, intelligence, and creative thinking in fifth-grade children: A correlational study. Merrill-Palmer Quarterly of Behavior and Development, 1966, 12, 233-241.

Footnotes

[1]This research was supported by Office of Economic Opportunity contract number OEO 4126.

[2]The authors would like especially to thank Norman Chansky for his continuing advice. Cathleen Kubiniec, Daniel Solomon, and James Roberge suggested helpful comments on a previous draft of this paper. Susan Lyons' coding and scoring help was most appreciated. Jeffrey O. Jones of the Philadelphia Get Set program willingly cooperated in this study.

[3]All children attending these centers were used except for a few who were termed untestable by their teachers for various reasons (speech problems, mental retardation, emotional disturbance). Subjects were then randomly selected from the eligible children.

Table 1

Variable		Trans-formation	Raw Score[a] Mean	SD
Construction Test	form A flexibility[b]	none	2.01	.82
Construction Test	form B flexibility	none	2.59	.89
Starkweather Test	form A fluency	normalized[c]	17.53	6.30
Starkweather Test	form B fluency	none	21.31	6.76
Starkweather Test	form A flexibility	normalized	15.76	5.19
Starkweather Test	form B flexibility	square root	18.43	5.49
Starkweather Test	form A originality	normalized	6.44	4.13
Starkweather Test	form B originality	square root	7.71	4.36
Pattern Meanings Test	form A fluency	normalized	9.59	4.19
Pattern Meanings Test	form B fluency	normalized	9.71	3.44
Pattern Meanings Test	form A flexibility	square root	8.86	3.16
Pattern Meanings Test	form B flexibility	square root	9.12	2.77
Pattern Meanings Test	form A originality	normalized	3.11	3.10
Pattern Meanings Test	form B originality	normalized	2.69	2.43

[a] The means and standard deviations are for an N of 175.

[b] All of the flexibility scores used in the analyses are the means of the two raters' scores.

[c] This transformation involves converting the score of every subject to a cumulative percentile rank. The percentile ranks are then transformed to z scores (Gulliksen, 1950).

Table 2

Product-Moment Intercorrelations of Creativity Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Construction Test-A-flexibility	--													
2. Construction Test-B-flexibility	.17	--												
3. Starkweather Test-A-fluency	.22	.14	--											
4. Starkweather Test-B-fluency	.13	.21	.55	--										
5. Starkweather Test-A-flexibility	.19	.15	.96	.54	--									
6. Starkweather Test-B-flexibility	.10	.26	.54	.96	.55	--								
7. Starkweather Test-A-originality	.24	.07	.83	.42	.79	.41	--							
8. Starkweather Test-B-originality	.14	.12	.51	.83	.48	.80	.45	--						

Table 2, Continued

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
9. Pattern Meanings Test-A-fluency	.05	.05	.13	.07	.11	.08	.01	.02	--					
10. Pattern Meanings Test-B-fluency	.09	.19	.04	.11	.04	.11	-.06	-.05	.25	--				
11. Pattern Meanings Test-A-flexibility	.05	.04	.17	.08	.15	.09	.02	.01	.93	.24	--			
12. Pattern Meanings Test-B-flexibility	.09	.15	.05	.12	.06	.11	-.05	-.01	.29	.90	.32	--		
13. Pattern Meanings Test-A-originality	.08	.00	.30	.22	.28	.23	.25	.22	.45	.05	.53	.14	--	
14. Pattern Meanings Test-B-originality	.07	-.10	.09	.10	.06	.06	.11	.15	-.01	.21	.03	.29	.23	--

$p < .05$ when $r = .15$ for $N = 175$ (two-tailed test)

$p < .01$ when $r = .20$ for $N = 175$ (two-tailed test)

Table 3

Split-half Reliability Correlations[a]

	Form A	Form B
Starkweather Test fluency	.84	.87
Starkweather Test flexibility	.82	.85
Starkweather Test originality	.72	.70
Pattern Meanings Test fluency	.81	.87
Pattern Meanings Test flexibility	.78	.79
Pattern Meanings Test originality	.62	.50

[a]N = 175. Corrected by the Spearman-Brown formula for a test of doubled length.

Table 4

Multivariate and Univariate Analysis of Variance of Form A of Creativity Tests

Source of Variation	Multivariate df	Multivariate F	Variable	Univariate df	Univariate F
Race	7/150	1.31	Construction Test flexibility	1/156	3.94*
			Starkweather Test fluency	1/156	2.42
			Starkweather Test flexibility	1/156	2.09
			Starkweather Test originality	1/156	3.06
			Pattern Meanings Test fluency	1/156	2.50
			Pattern Meanings Test flexibility	1/156	1.42
			Pattern Meanings Test originality	1/156	.00
Sex	7/150	.72	Construction Test flexibility	1/156	.08
			Starkweather Test fluency	1/156	.55
			Starkweather Test flexibility	1/156	1.16
			Starkweather Test originality	1/156	.55
			Pattern Meanings Test fluency	1/156	.59
			Pattern Meanings Test flexibility	1/156	.60
			Pattern Meanings Test originality	1/156	2.55
Test Condition	14/300	.70	Construction Test flexibility	2/156	.81
			Starkweather Test fluency	2/156	.48
			Starkweather Test flexibility	2/156	.73
			Starkweather Test originality	2/156	.77
			Pattern Meanings Test fluency	2/156	1.14
			Pattern Meanings Test flexibility	2/156	1.17
			Pattern Meanings Test originality	2/156	1.14

Table 4, Continued

Source of Variation	Multivariate df	Multivariate F	Variable	Univariate df	Univariate F
Race x Sex	7/150	.54	Construction Test flexibility	1/156	.22
			Starkweather Test fluency	1/156	.36
			Starkweather Test flexibility	1/156	.08
			Starkweather Test originality	1/156	.30
			Pattern Meanings Test fluency	1/156	.02
			Pattern Meanings Test flexibility	1/156	.08
			Pattern Meanings Test originality	1/156	1.54
Race x Test Condition	14/300	.84	Construction Test flexibility	2/156	.01
			Starkweather Test fluency	2/156	.36
			Starkweather Test flexibility	2/156	.20
			Starkweather Test originality	2/156	.20
			Pattern Meanings Test fluency	2/156	.47
			Pattern Meanings Test flexibility	2/156	.95
			Pattern Meanings Test originality	2/156	.60
Sex x Test Condition	14/300	.47	Construction Test flexibility	2/156	.61
			Starkweather Test fluency	2/156	1.65
			Starkweather Test flexibility	2/156	1.68
			Starkweather Test originality	2/156	1.47
			Pattern Meanings Test fluency	2/156	.45
			Pattern Meanings Test flexibility	2/156	.66
			Pattern Meanings Test originality	2/156	.49

Table 4, Continued

Source of Variation	Multivariate df	Multivariate F	Variable	Univariate df	Univariate F
Race x Sex x Test Condition	14/300	1.07	Construction Test flexibility	2/156	3.12*
			Starkweather Test fluency	2/156	1.07
			Starkweather Test flexibility	2/156	1.15
			Starkweather Test originality	2/156	3.36*
			Pattern Meanings Test fluency	2/156	1.04
			Pattern Meanings Test flexibility	2/156	1.31
			Pattern Meanings Test originality	2/156	.22

* $p < .05$

Table 5

Multivariate and Univariate Analysis of Variance of Form B of Creativity Tests

Source of Variation	Multivariate df	F	Variable	Univariate df	F
Race	7/150	.65	Construction Test flexibility	1/156	2.33
			Starkweather Test fluency	1/156	.81
			Starkweather Test flexibility	1/156	1.10
			Starkweather Test originality	1/156	.07
			Pattern Meanings Test fluency	1/156	.22
			Pattern Meanings Test flexibility	1/156	.49
			Pattern Meanings Test originality	1/156	.70
Sex	7/150	.68	Construction Test flexibility	1/156	.00
			Starkweather Test fluency	1/156	1.43
			Starkweather Test flexibility	1/156	1.00
			Starkweather Test originality	1/156	1.04
			Pattern Meanings Test fluency	1/156	.38
			Pattern Meanings Test flexibility	1/156	.01
			Pattern Meanings Test originality	1/156	.30
Test Condition	14/300	.87	Construction Test flexibility	2/156	.16
			Starkweather Test fluency	2/156	2.44
			Starkweather Test flexibility	2/156	2.55
			Starkweather Test originality	2/156	.52
			Pattern Meanings Test fluency	2/156	.38
			Pattern Meanings Test flexibility	2/156	.19
			Pattern Meanings Test originality	2/156	.92

Table 5, Continued

Source of Variation	Multivariate df	Multivariate F	Variable	Univariate df	Univariate F
Race x Sex	7/150	.52	Construction Test flexibility	1/156	1.22
			Starkweather Test fluency	1/156	.18
			Starkweather Test flexibility	1/156	.62
			Starkweather Test originality	1/156	.04
			Pattern Meanings Test fluency	1/156	.02
			Pattern Meanings Test flexibility	1/156	.22
			Pattern Meanings Test originality	1/156	.01
Race x Test Condition	14/300	.53	Construction Test flexibility	2/156	.11
			Starkweather Test fluency	2/156	.71
			Starkweather Test flexibility	2/156	.91
			Starkweather Test originality	2/156	.85
			Pattern Meanings Test fluency	2/156	.23
			Pattern Meanings Test flexibility	2/156	.22
			Pattern Meanings Test originality	2/156	.12
Sex x Test Condition	14/300	2.47**	Construction Test flexibility	2/156	.88
			Starkweather Test fluency	2/156	.22
			Starkweather Test flexibility	2/156	1.36
			Starkweather Test originality	2/156	.48
			Pattern Meanings Test fluency	2/156	3.23*
			Pattern Meanings Test flexibility	2/156	2.09
			Pattern Meanings Test originality	2/156	2.34

Table 5, Continued

Source of Variation	Multivariate df	Multivariate F	Variable	Univariate df	Univariate F
Race x Sex x Test Condition	14/300	1.09	Construction Test flexibility	2/156	.09
			Starkweather Test fluency	2/156	2.08
			Starkweather Test flexibility	2/156	1.76
			Starkweather Test originality	2/156	1.29
			Pattern Meanings Test fluency	2/156	.27
			Pattern Meanings Test flexibility	2/156	.02
			Pattern Meanings Test originality	2/156	1.05

* $p < .05$

** $p < .003$

Table 6

Raw Scores for the Pattern Meanings Test, Fluency Score, Form B

	Boys		Girls	
	Mean	SD	Mean	SD
Play	10.14	2.35	8.93	4.20
Verbal	9.32	2.13	10.61	3.81
Nonverbal	9.75	3.23	9.89	4.07

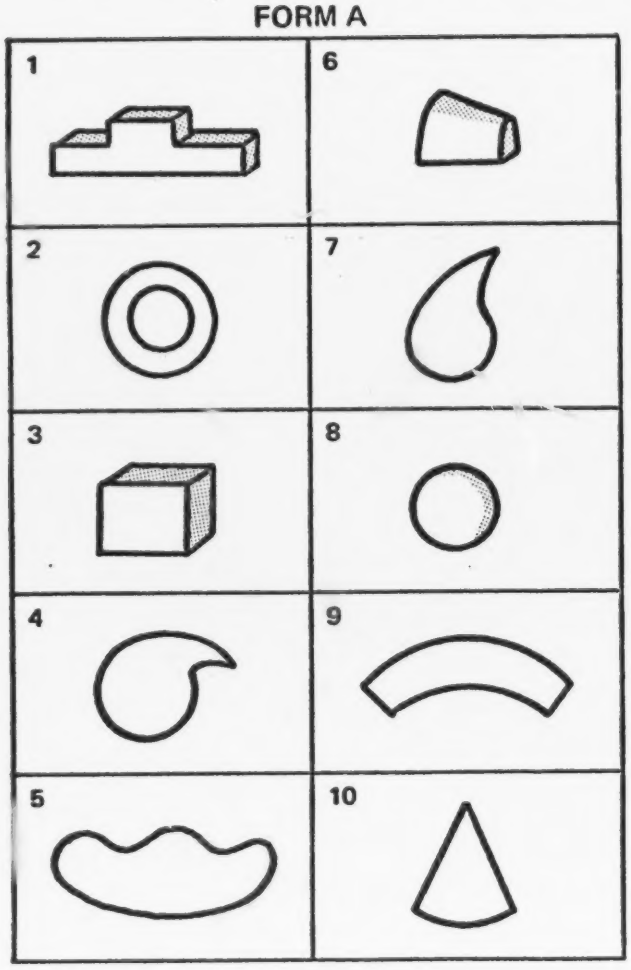
N = 28 subjects/cell

Figure 1

Starkweather Test Styrofoam Forms

STARKWEATHER TEST

FORM A



FORM B

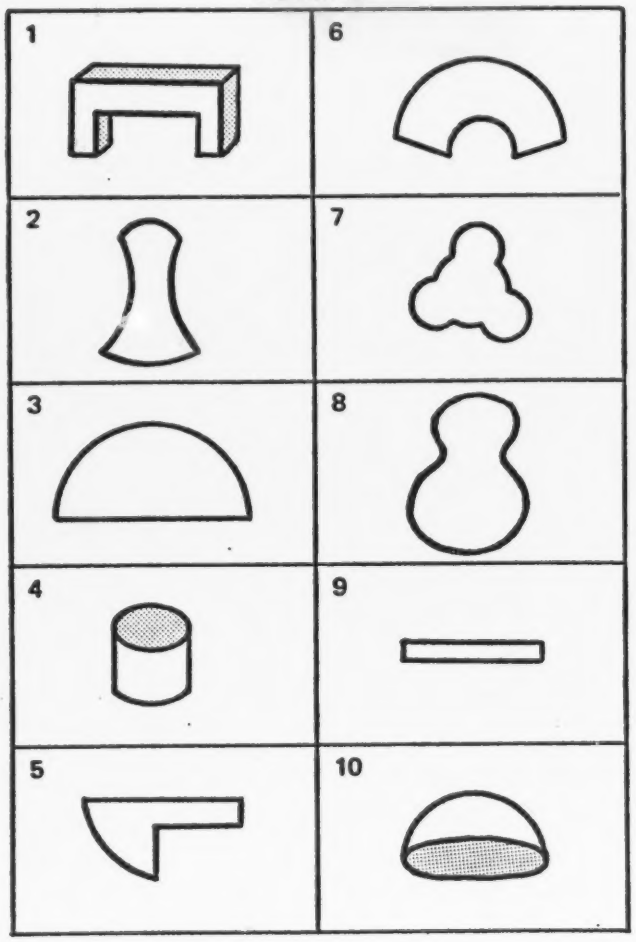


Figure 2

Pattern Meanings Test Figures, Form B

PATTERN MEANINGS TEST: FORM B